

12

AD-A257 363



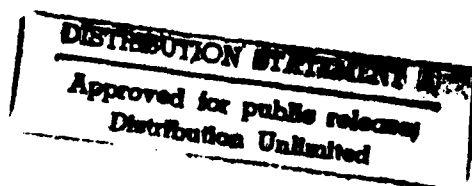
ADAPTIVE MACHINE VISION
FINAL REPORT

ONR Contract N00014-86-C-0601

A. Evan Iverson
William W. Stoner



SCIENCE APPLICATIONS INTERNATIONAL CORPORATION
6 Fortune Drive
Billerica, Massachusetts 01821
(508) 667-6365



92-29078



20PY

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Adaptive Machine Vision		5. TYPE OF REPORT & PERIOD COVERED final report 10/86-9/92
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A. Evan Iverson and William W. Stoner		8. CONTRACT OR GRANT NUMBER(s) N00014-86-C-0601
9. PERFORMING ORGANIZATION NAME AND ADDRESS SCIENCE APPLICATIONS INTERNATIONAL CORPORATION 6 Fortune Drive Billerica, MA 01821		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Strategic Defense Initiative Organization Innovative Science and Technology Office, Washington, D. C.		12. REPORT DATE 3 November 1992
		13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Machine Vision, Pattern Recognition		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The final report summarizes the work on multi-target tracking, adaptive photodetector array concepts for sensors that must operate in nuclear disturbed backgrounds, pattern detection for target/decoy discrimination with a neural network, and Gabor wavelet techniques for feature extraction in a neural network inspired by the Neocognitron of Fukushima.		

1.0 Overview

Since the Contract start in October 1986, progress was made in several areas. In this report we summarize our progress and report on work in the last phase of the Contract. Our most recent work concerns the application of Gabor Wavelets to feature extraction and pattern recognition for postboost or midcourse phase ballistic missile defense. From the point of view of an interceptor, a target changes in scale rapidly during the last seconds before impact. Pattern recognition during these last seconds is needed for terminal guidance of the interceptor. Effective pattern recognition across a large range of scale size requires scale invariant feature extraction. This requirement motivates a wavelet representation of the input scene. An infinite number of possible wavelet representations may be considered, but Gabor wavelets are the obvious candidates for optical information processing. The subject of wavelet representations is covered in Section 3.0. A summary of progress made in other areas is given below for each annual reporting period of the Contract.

1.1 Period from October 1986 to Dec 1988

March 1988 Annual report: Precision metrics on a ballistic target are obtained by following the target over a period of time, with either angle measurements, range measurements or both angles and range. These measurements are processed with a batch least-squares estimator to obtain the refined metric information. In a multi-target environment, targets A and B may fall in the same resolution cell of the framing camera or pulsed radar. Once this happens, the identities of the two targets are ambiguous. How do we assign earlier measurements to a target after the two targets cross in angle or range? If we attempt to sort out this confusion by trying all possible assignments of the measurements, we face a very large computational load, because the number of ways to combine measurements from two or more targets grows very rapidly with the number of measurements used in the batch least-squares estimator. This data association problem occurs in its simplest form as frame-to-frame correlation: the pairing up of target blips appearing on consecutive image frames.

A related problem is encountered in matching stereo pairs for machine vision, where it is called the "correspondance problem." A possible solution to the correspondance problem proposed by Professor Eric Schwartz of NYU was studied because it lends itself to

DISQUALITY INSPECTED

For	
AI	<input checked="" type="checkbox"/>
ad	<input type="checkbox"/>
ion	<input type="checkbox"/>
on/	
Availability Codes	
Dist	Avail and/or
A-1	Special

an optical implementation. Schwartz proposes that disparity between stereo pairs might be extracted automatically by cepstral processing. The cepstrum was proposed by J. W. Tukey for detection of echos in signals. Since one of the stereo pairs may be regarded as the "echo" of the other, where the echo delay is the stereo disparity, cepstral processing also applies to the problem of finding corresponding features in stereo pairs, and extracting the stereo disparity to find the range of a feature located in both stereo pairs.

We also studied the computational complexity of the Neocognitron. The Neocognitron is a neural network developed by Fukushima and his colleagues (K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by a shift in position," Biological Cybernetics, vol. 36, pp 193-202, 1980). The self-organizing characteristic of the Neocognitron refers to its ability to group input scenes into categories without a teacher identifying the categories during training. The mechanism behind this "learning without a teacher," is known as "competitive learning." Optical processing does not lend itself to implementing competitive learning. A hybrid system inspired by the Neocognitron is now being investigated by Dr. T. H. Chao at the Jet Propulsion Laboratory in Pasadena CA. An optical correlator performs the feed-forward action of the Neocognitron, and a digital computer carries out the learning mechanism and changes the filters in the optical correlator. Our work in collaboration with Dr. T. H. Chao has been supported by the Contract since June 1990 (publications 4, 5, 6 and 7 listed in Section 1.7).

Since SDI sensors must operate in a nuclear disturbed environment with striated brightness variations in the foreground or background, automatic compensation within the photosensitive detector array for local light level is important. This concern motivates our investigation of a push-pull mechanism for adaptive spatial resolution and image contrast developed by Dr. M. H. Brill. We performed computer simulations on the push-pull mechanism, and confirmed its adaptive spatial resolution and adaptive image contrast properties.

1.2 Period from Jan 1988 to Dec 1988

December 1988 Annual Report: We continued our work on the adaptive photosensor array model and the correspondance problem in stereo vision. We generated three publications on the adaptive

photosensor array model (publications 1-3 listed below in Section 1.7).

1.3 Period from Dec 1988 to Jan 1989

Report dated Jan 1989: progress in several areas are covered.

1) We continued work on the adaptive photosensor array and looked into hardware implementations.

2) We performed lengthy simulations of the performance of several multi-target tracking algorithms that lend themselves to optical processing. We investigated the well-known Hough transform method, as well as novel methods we devised to improve performance over the Hough transform. These novel methods are the "isometric projection" method and the "3-D track search" method. The weakness of the Hough transform in multi-target tracking is that the time sequencing of the observations along a trajectory is lost. This is a considerable disadvantage for multi-target tracking, because targets that do not simultaneously occupy nearby points in space are lumped together in the Hough transformation. This lumping of events (that are distinguishable if the temporal sequencing is taken into account) makes the multi-target problem much worse than it really is. Using a Relative Operating Characteristic (ROC) formalism to rate the three methods, we found the isometric projection method and the 3-D track search method provide significantly better performance than the Hough transform.

3) Neural network receptive fields for uniform weighting of an input field were investigated as part of our continuing work on the Neocognitron. In the Neocognitron, shift invariant pattern recognition is sought by massive spatial replication of the feature detectors. However, unless the receptive fields of the feature detectors are overlapped very carefully, the output of the Neocognitron will be sensitive to a shift of the input pattern. We investigated this problem for 1-D and 2-D input windows, and devised a solution to the problem by appropriately shaping the receptive fields.

4) A scanning sensor may be required to cover the large field-of-views encountered in Ballistic Missile Defense (BMD). Such a scanning sensor generally uses a pushbroom scan of a linear array detector to sweep out a 2-D field-of-view over each frame period.

Examination of this scanning scheme shows that it does not satisfy the Nyquist sampling theorem. We show how to satisfy the Nyquist sampling criterion with a scanning linear array detector by staggering the detector elements in the array.

5) Animal vision systems frequently use hexagonal sampling patterns instead of the square sampling patterns generally in use in electronic cameras. We derive a sampling theorem for hexagonal sampling and show that hexagonal sampling is slightly more efficient than square sampling.

1.4 Funding hiatus Jan 1989 to June 1990 (no reports generated).

1.5 Period from June 1990 to December 1991

Report dated June 1992: With the resumption of funding, we began to work closely with T. H. Chao at the Jet Propulsion Laboratory in Pasadena California. Dr. Chao is funded to develop an optical neural network for SDI target/decoy discrimination. We worked in support of this goal by:

- 1) helping to structure the architecture of the optical neural network,
- 2) performing simulations of target and decoy images for development and testing of the optical neural network, and
- 3) developing a mathematical basis for simulating optical images of objects modelled with constructive solid geometry.

We reported on this work in a poster paper at the Gordon conference on Coherent Optics and Holography, Plymouth State College, NH, June 1991, and at the July 1991 SPIE meeting in San Diego CA (publications 4-5 in Section 1.7 below).

1.6 Period from Jan 1991 to Sept 1992

Final Report: during 1992, we concentrated on improving the architecture of the optical neural network under development at JPL by Dr. T. H. Chao. This optical neural network uses an optical correlator for feature extraction. Since the inspiration for the neural network architecture is the Neocognitron of Fukushima, we originally copied the feature detectors used in some of Fukushima's papers for

the first stage of feature extraction. Fukushima designed the Neocognitron for recognition of handwritten alphanumeric characters. Consequently, at the first stage of feature extraction, he was able to use line segments at different orientations to characterize the individual pen strokes in alphanumeric characters. However, our application to SDI target/decoy discrimination deals with continuous-tone images, not line figures. We could force the continuous-tone input images to be line figures by edge-enhancement, but this would throw away shading information. Since shading information can be used to infer shape, it seems best to preserve it. We conclude that the line segment features used successfully by Fukushima are not appropriate for our application.

We seek general features, so that relevant image information is not lost in the initial stage of processing. The Gabor wavelets are a natural choice, because they extract enough information to accurately reproduce the input, and they provide a uniform set of features across a wide range of scales. Section 2.0 discusses the BMD problem and motivates the application of wavelets to the target/decoy discrimination problem. Section 3.0 provides more information on wavelet theory for feature extraction.

1.7 Publications And Presentations

- 1) M. H. Brill, D. Bergeron and W. W. Stoner, "Retinal Model with Adaptive Contrast and Resolution," *Applied Optics*, pp. 4993-4998, 1 Dec 1987.
- 2) M. H. Brill, D. Bergeron and W. W. Stoner, "Trichromatic Retinal Model with Adaptive Contrast Sensitivity and Resolution," talk given at the Annual Optical Society of America Meeting, Rochester, NY, 22 Oct 87.
- 3) Michael H. Brill, Doreen W. Bergeron and William W. Stoner, "A Model Retina with Push-Pull Cooperative Receptors Providing Adaptive Contrast Sensitivity and Resolution," pp. 101-106 of the DARPA Neural Network Study, issued 18 July 1988.
- 4) Tein-Hsin Chao, H. Langenbacher, Sam Rosenzweig, and W. W. Stoner, "Radar Discrimination with an Optical Neural Network (RADONN)," poster paper presented at the Gordon Research

Conference on Holography and Information Processing, 17-21 June 1991, Plymouth State College, Plymouth NH.

5) Tien-Hsin Chao and William W. Stoner, "Optical Implementation of Neocognitron and its applications to radar signature discrimination," Proceedings of SPIE Vol. 1558, Wave Propagation and scattering in varied media II, pp. 505-517, July, 1991.

6) Tien-Hsin Chao, William J. Miceli, and William W. Stoner, "Optical Implementation of a shift-invariant Neocognitron," Proceedings of SPIE Vol. 1703, in press. Presented at SPIE meeting, San Diego CA, July, 1992.

7) Tien-Hsin Chao and William W. Stoner, "Optical implementation of a feature-based neural network with application to automatic target recognition," accepted for publication by Applied Optics special issue on optical implementation of neural networks, edited by D. Psaltis and K. Wagner (to appear in March, 1993).

2.0 The SDI Decoy Discrimination Problem

In this section, an overview of the quintessential SDI problem of ballistic missile defense is described in terms of the four phases of payload delivery. The use of decoys in the mid-course phase is then discussed and an example of the magnitude of the mid-course defense problem is presented. Finally, aspects of the mid-course decoy discrimination problem pertaining to imaging sensors are presented.

2.1 The Ballistic Missile Defense Problem

The delivery of a weapon payload by an intercontinental ballistic missile (ICBM) can be divided into four phases: boost, post-boost, mid-course, and reentry. All phases offer opportunities for defensive action, however, the approach and required technology vary considerably for each phase.

The boost phase begins at launch and ends approximately three to six minutes later with the termination of thrust at an altitude of 200 to 300 km. By the end of the boost phase, the lift vehicle has given the payload sufficient velocity to reach the target. Ballistic missile defense (BMD) during the boost phase relies on infrared (IR) emission from the rocket plume as the most prominent observable to use for the defensive system. Boost-phase BMD offers high leverage because of the high value of the target, however, the time available for target acquisition, targeting, and destruction is relatively short. Furthermore, the development of fast burn boosters may potentially reduce the boost phase to approximately one minute.

Following the boost phase, the elements of the payload are separated from the lift vehicle and fall in ballistic trajectories until impact. In multiple, independently targeted reentry vehicle (MIRV) systems, small velocity increments are given to each reentry vehicle (RV) to direct them towards individual targets. During this post-boost phase, a missile stage known as the post-boost vehicle or *bus* is utilized to deploy the RVs along with defenses such as decoys or other penetration aids. At preprogrammed positions and velocities, the bus releases single RVs and/or multiple decoys. Currently, from 4 to 10 RVs per missile can be lifted with the bulk of the Russian land-based strategic ballistic missiles which consists of SS-17, SS-18, and SS-19 ICBMs.

The post-boost phase lasts approximately 5 minutes and the observables are much weaker than in the boost phase. Three candidate sensor technologies are apparent for post-boost defense, microwave radar, laser radar, and passive thermal detection. Rapid post-boost defense is highly desirable because the number of potential targets to be tracked becomes considerably greater as the RVs and decoys are deployed.

Following the boost and post-boost phases, surviving elements of the offense enter the mid-course phase. For an ICBM or SLBM, the mid-course phase is the longest of the trajectory phases (approximately 20 minutes for intercontinental ranges). Throughout the mid-course phase, all RVs and decoys from a given missile, as well as remnants of the bus and booster, move along adjacent ballistic trajectories under the influence of gravity. Mid-course countermeasures to BMD are made easier by the lack of atmosphere which allows lightweight decoys to travel in the same trajectories as the heavy RVs. The principal problem to be confronted in mid-course BMD is the discrimination of RVs from decoys in a heavy traffic environment. These decoys may be close replicas of the RVs or they may be simpler *traffic* decoys that are deployed in large numbers to saturate the defensive sensors. The problem of discriminating these decoys from the RVs provides the motivation for this report and is discussed further in Sections 2.2 and 2.3.

The reentry phase begins when the threat cloud begins to enter the earth's atmosphere. During reentry, the difficult discrimination problems associated with mid-course defense disappear or are greatly reduced. Changes in the trajectories of the lightweight decoys make them increasingly easy to discriminate from RVs. The optical signature of reentering bodies is also increased due to frictional heating. These tracking and discrimination opportunities are, however, largely offset by the short time available (less than 60 seconds) for terminal defense.

2.2 Decoys and the Mid-Course BMD Problem

The use of decoys and penetration aids in the mid-course phase offers an effective BMD countermeasure for the offense. This advantage is partially offset by the relatively long time the defense has to establish trajectories and perform discrimination. It is well known that the mid-course BMD problem is critically dependent on the success of boost-phase and post-boost defense. A simple

example servers to illustrate the magnitude of the mid-course BMD problem in terms of the number of objects to be acquired, tracked, discriminated, and destroyed.

Preliminary designs suggest that effective replica decoys, having the same shape and size of an RV, can be constructed with a mass of 1-2 kg. An RV with a 200-kt warhead weighs approximately 200 kg, therefore, 100 to 200 decoys can be deployed at the cost of offloading a single RV in a MIRV system. Consider a typical booster with a payload capacity of 4000 kg. Using 2-kg decoys, a single booster can potentially deliver 10 RVs and 1000 decoys into the mid-course battle. If the boost and post-boost defenses allow the payloads of only 100 missiles to survive, the mid-course defense system can be faced with the problem of acquiring, tracking, and discriminating 1000 RVs interspersed among 100,000 decoys. All this plus the destruction of the RVs must be performed in less than one-half hour. If the boost and post-boost defenses are one tenth as effective, mid-course defense can be faced with the massive problem of destroying 10,000 RVs interspersed among 1,000,000 decoys. The magnitude of the task to be performed by mid-course defense is thus critically dependent on the success of boost and post-boost defense.

2.3 The Decoy Discrimination Problem

It is unlikely that observation of boost-phase deployment can provide high-confidence discrimination between RVs and decoys, therefore, the defensive system must acquire information during the mid-course phase that allows discrimination. The offensive strategy is to minimize or mask the signatures available to perform this discrimination. Characteristics available for discrimination fall into two categories: (1) *intrinsic* characteristics, such as mass and the presence of fissile materials, high explosives, electronics, etc.; and (2) *extrinsic* characteristics, such as size, shape, surface temperature, and surface emissivity.

Intrinsic characteristics are difficult or impossible to ascertain. Without thrust and drag, RVs and decoys move only under the influence of gravity, therefore, it is impossible to discriminate by mass through metric tracking. *Interactive perturbing techniques* that deliver an impulsive force on the objects in the threat cloud are required for discrimination based on mass. Detecting the presence of fissile materials or explosives requires a highly directive, penetrating sensor capable of discrimination based on nuclear or chemical

signatures. Electronics may be detectable by electromagnetic emissions, however, jamming and the use of replica electronics on the decoys reduces the utility of electromagnetic emissions for discrimination.

Extrinsic characteristics offer signatures for discrimination that can be obtained by direct observation with imaging or spectroscopic sensors. In this report, we focus on the problem of performing imagery-based decoy discrimination. Passive imaging sensors such as long-wavelength infrared (LWIR) and active imaging sensors such as microwave radar and laser radar are the most likely candidates for detection based on extrinsic characteristics.

Decoys for the exoatmospheric mid-course phase are typically lightweight objects inflated or erected in space. Some decoys may also be designed to function into the early part of the reentry phase, therefore, they will require heavier designs in order to duplicate the aerodynamic properties of real RVs. Ideally the decoys would have size, shape, and surface characteristics matching those of an RV. In practice, this is not possible. To provide decoys that are compact to store in the bus, easy to deploy, and lightweight limits the accuracy to which a decoy can match the characteristics of an RV.

The characteristics of the decoy are selected by the offense to minimize the difference in signature between a decoy and an RV for a given sensor or suite of sensors. These differences must be fully exploited by the defensive system. The problem of discriminating these decoys from the RVs using LWIR or radar imagery requires advanced techniques and algorithms for processing this imagery. In Section 3.0 of this report, we present a technology that is directly relevant to this particularly challenging problem.

3.0 Wavelet Multiresolution Analysis for Imagery-Based Decoy Discrimination

The basic problem of imagery-based decoy discrimination falls in the area of pattern recognition and object classification. This area has evolved considerably over the past 20 years. During that time, much progress has been made in the development of approaches for situations in which image formation conditions, such as distance, orientation, background, and illumination can be carefully controlled. The problem of decoy discrimination based on extrinsic characteristics, such as size and shape, relies on obtaining imagery

using either passive or active sensors under conditions which do not allow such careful control of the imaging process. Fortunately, the range of conditions for a particular sensor can be predicted quite well prior to the development of the object classification algorithm.

One critical aspect of the image-based decoy discrimination problem is that of addressing differences in orientation and spatial scale of objects in the imagery. Because of these factors, techniques for the extraction and representation of object features required for object classification must appropriately designed. In this section, we present feature-extraction and image-representation techniques that are potentially significance for the problem of imagery-based decoy discrimination. In Section 3.1, we present an overview of joint spatial/spatial-frequency representations as they apply to feature extraction. This overview also provides context for the discussion of wavelets. In Section 3.2, the the subject of image resolution pyramids and subband coding is presented as an introduction to the concept of multiresolution representations. Finally, in Section 3.3, the subject of wavelets and wavelet multiresolution analysis is presented. Wavelet-based techniques are seen to offer potential solutions to problems associated with differences in orientation and spatial scale.

3.1 Joint Spatial/Spatial-Frequency Representations

One approach to the discrimination of decoys from RVs in imagery requires the extraction of object features in order to perform object classification. The use of both spatial information and spatial-frequency information in a unified manner provides a considerable advantage in feature-based object classification. Joint spatial/spatial-frequency methods are based on image representations that give the spatial frequency content in localized regions in the spatial domain. These methods overcome many of the shortcomings of traditional Fourier transform methods. They can also provide balanced resolution in both domains and are consistent with known characteristics of the human visual system.

In this section, we present an overview of some of the classical joint spatial/spatial-frequency representations that have utility as feature measures. These representations we present are (1) the short-time Fourier transform, (2) the Gabor representation, and (3) the Wigner distribution. The presentation of these methods also provides context for the wavelet material presented in Section 3.3.

3.1.1 The Short-Time Fourier Transform and the Spectrogram

The one-dimensional *widowed* or *short-time* Fourier transform (STFT) has been used extensively in the analysis of temporal signals. For temporal signals, the STFT produces a two-dimensional time-frequency representation. This concept is easily extended to the spatial image domain to yield a four-dimensional spatial/spatial frequency representation. This result, known as the finite-support windowed Fourier transform, can be written in continuous variables as:

$$F(x', y', u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot g(x - x', y - y') \cdot e^{-i(u x + v y)} dx dy$$

where $f(x, y)$ is the spatial image data and $g(x - x', y - y')$ is a window function centered at position (x', y') .

The spectrogram, defined as the squared magnitude of $F(x', y', u, v)$, has been used by Bajcsy and Lieberman [1] as a texture feature measure and by Pentland [2] for estimating fractal dimension.

3.1.2 The Gabor Representation

The complex, two-dimensional Gabor functions have the general form:

$$h(x, y) = g(x', y') \cdot \exp[2\pi i(Ux + Vy)]$$

where $(x', y') = (x \cos \phi + y \sin \phi, -x \sin \phi + y \cos \phi)$ are rotated coordinates and where

$$g(x, y) = \left(\frac{1}{2\pi\lambda\sigma^2} \right) \cdot \exp \left[-\frac{(x/\lambda)^2 + y^2}{2\sigma^2} \right].$$

This function can be thought of as a complex sinusoidal grating modulated by a two-dimensional Gaussian function with aspect ratio λ , scale parameter σ , and major axis oriented at angle ϕ from the x -axis. If $\lambda = 1$, then ϕ need not be specified since $g(x, y)$ is circularly symmetric. The spatial frequency response of the Gabor function is

$$H(u, v) = \exp \left\{ -2\pi^2 \sigma^2 \left[(u' - U)^2 \lambda^2 + (v' - V)^2 \right] \right\}$$

where $(u', v') = (u \cos \phi + v \sin \phi, -u \sin \phi + v \cos \phi)$ and (U, V) is a similar rotation of center frequency (U, V) . Thus, $H(u, v)$ is a bandpass Gaussian with minor axis oriented at an angle ϕ from the u -axis, aspect ratio $1/\lambda$, radial center frequency $F = \sqrt{U^2 + V^2}$ (measured in cycles/image), and orientation $\theta = \tan^{-1}(V/U)$ measured from the u -axis.

Any finite-dimensional function can be expressed as a weighted sum of appropriately shifted Gabor functions. This sum is known as the Gabor representation. If $f(x,y)$ is the spatial image data, then it can be represented with the sum:

$$f(x,y) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \beta_{rsmn} g(x-x_r, y-y_s) \cdot \exp\{2\pi i[U_m(x-x_r) + V_n(y-y_s)]\}$$

where the sequences of shifts $\{x_r\}$ and $\{y_s\}$ and modulation frequencies $\{U_m\}$ and $\{V_n\}$ have constant spacings X, Y, U , and V satisfying $XU = YV = 1$. The resulting grid of shifts and frequencies (four-dimensional in this case) is known as the Gabor lattice. The expansion coefficients $\{\beta_{rsmn}\}$ form a complete representation of $f(x,y)$ since $f(x,y)$ can be exactly reconstructed from the coefficients.

Recently, Porat and Zeevi [3] have proposed a generalized Gabor representation for use in machine vision. Clark *et. al.* [4] have used a set of Gabor filters to investigate texture segmentation.

3.1.3 The Theory of Time-Frequency Distributions and the Wigner Distribution

The concept of a mathematical representation in the form of a joint distribution function of time and frequency was motivated by the need for a rigorous mathematical analysis and clarification of the concept of time/frequency analysis. Beginning with the classical works of Gabor [5], Ville [6] and Page [7], the mathematically rigorous study of time-varying spectra was developed. The basic concept is to devise a joint distribution function of time and frequency that describes the energy density of a signal simultaneously in the time and frequency domain in a manner similar to the STFT.

There are many (in fact infinitely many) different possible time frequency distributions. The problem is to find a distribution that gives results consistent with several intuitive criteria. Different distribution arise from applying different criteria. The general form of the time-frequency distribution [8] is

$$P(\tau, \omega) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z\left(u + \frac{w}{2}\right) z^*\left(u - \frac{w}{2}\right) \phi(v, w) e^{-i(v\tau + w\omega - vw)} du dv dw$$

where $\phi(v, w)$ is an arbitrary kernel function and $z(t)$ is the analytic signal associated with the real signal to be analyzed, $f(t)$. The

analytic signal is calculated, using the quadrature function produced by the Hilbert transform, as

$$f(t) = s(t) + iH[f(t)](t)$$

where $H[\bullet]$ denotes the Hilbert transform defined as

$$H[f(t)](t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{f(\xi)}{\xi - t} d\xi,$$

and where P indicates that the Cauchy principal value of the integral is to be taken.

Different distributions are obtained for different kernel functions $\phi(v, w)$. The choice of kernel function can provide behavior that satisfies some of the intuitive criteria, although no distribution has yet been found that satisfies all of the proposed criteria. The simplest, and currently most common, distribution is the Wigner (or Wigner-Ville) distribution (WD) [8,9] given by the kernel $\phi(v, w) \equiv 1$. Carrying out the integration with respect to v and w , the Wigner distribution $W_f(\tau, \omega)$ of the function $f(t)$ can be written as

$$W_f(\tau, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} z\left(\tau + \frac{u}{2}\right) z^*\left(\tau - \frac{u}{2}\right) e^{-i\omega u} du.$$

The WD has the property that, for each time-shift value τ , the frequency direction centroid is the derivative of phase (instantaneous frequency) of the analytic signal. The principal advantage of the WD over the STFT is that the WD gives better performance on non-stationary functions. The WD, however, has the disadvantage of cross-product interference. The Choi-Williams distribution [10] represents a potential advantage with regard to the problem of cross-product interference.

The Wigner distribution was originally suggested to characterize the quantum-mechanical duality between the position and momentum of a particle [9]. Later, the WD was used by J. Ville for signal analysis [6], hence the frequent reference as the Wigner-Ville distribution in the signal processing literature. An excellent series of articles [11, 12, 13] have been published on the WD that presents the properties of the WD as well as examples of the WD applied to functions of one variable. The one-dimensional WD has been applied in a number of areas including speech analysis and optics. The use of the WD for 2-D and 3-D image processing was first advanced by Jacobson and Wechsler [14,15]. In the case of two-dimensional image data, the WD of the continuous image $f(x, y)$ is defined as

$$W_f(x, y, u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_f(x, y, \alpha, \beta) \cdot e^{-i(\alpha u - \beta v)} d\alpha d\beta$$

where $R_f(x, y, \alpha, \beta) = f\left(x + \frac{\alpha}{2}, y + \frac{\beta}{2}\right) f^*\left(x - \frac{\alpha}{2}, y - \frac{\beta}{2}\right)$. In this case, the extension to an analytic function is not performed.

The two-dimensional WD has a property that is of particular interest in image processing, namely, it is strictly a real-valued function. This property implies that the WD lacks the explicit phase component that is present in the STFT or Fourier transform. Nonetheless, an image can be completely recovered (to within a minus sign) from its Wigner distribution. This indicates that phase information is implicit in the WD. Discrete computational approximations to the WD have been defined by Reed and Wechsler [16].

3.2 Image Resolution Pyramids and Subband Coding

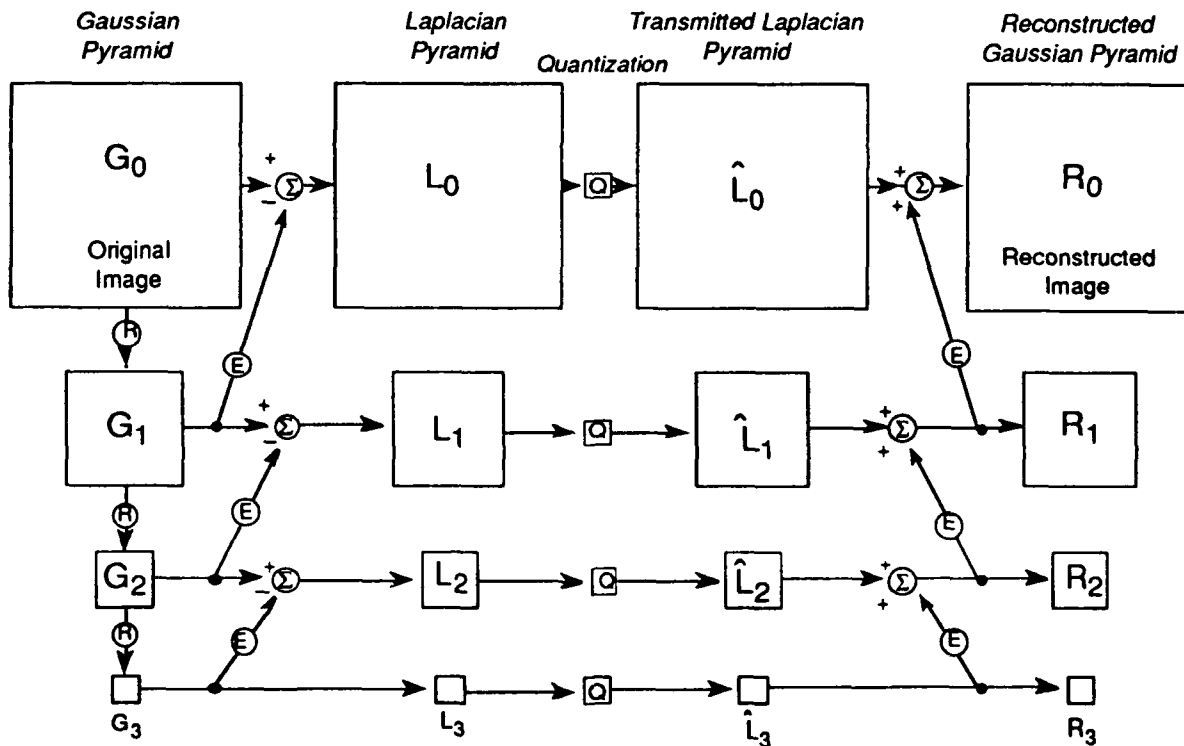


Figure 1. Block diagram of the Laplacian Pyramid of Burt and Adelson

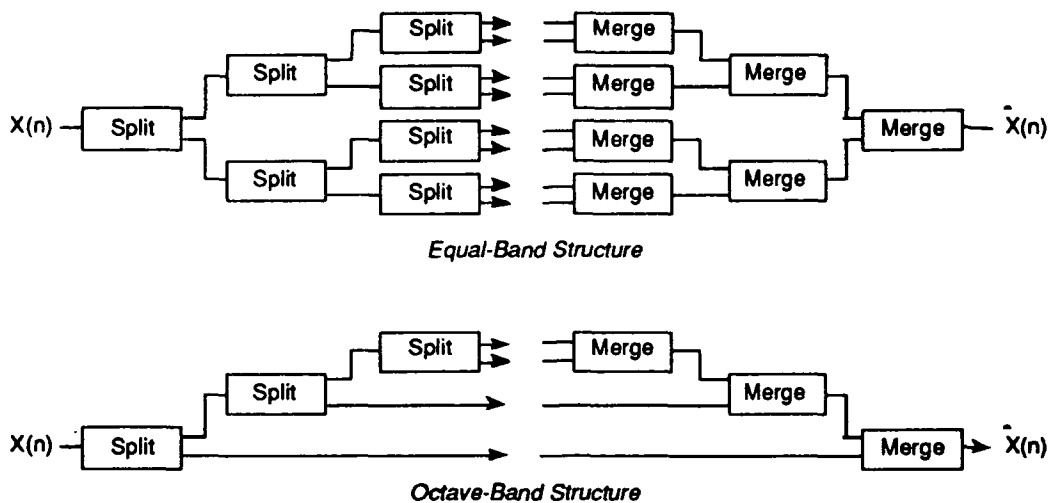


Figure 2. Block diagram of two commonly used subband structures for one-dimensional data.

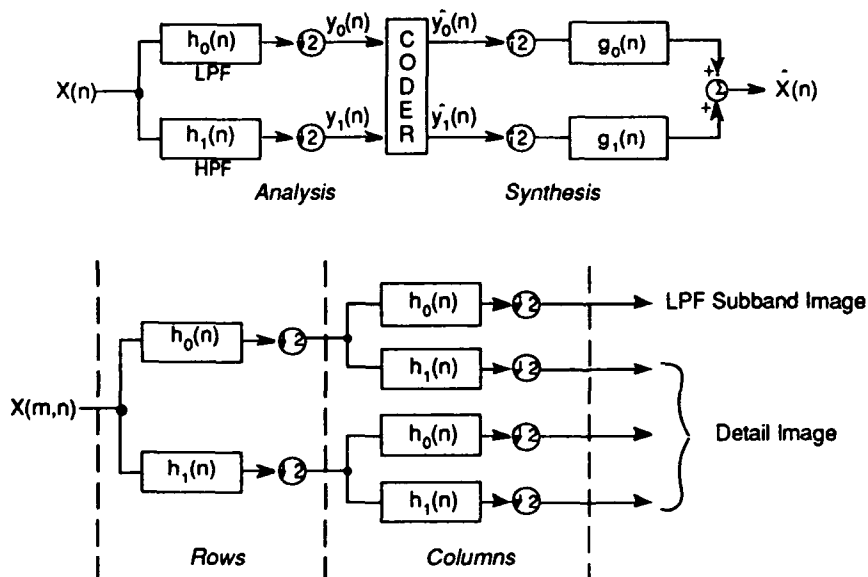


Figure 3. Block daigram of the two-band subband coder for one-dimensional data (top) and the seperable extension to two-dimensional data (bottom).

3.3 Wavelets and Multiresolution Analysis

Advances in signal and image processing have recently been brought about by the discovery of *compactly supported wavelets* and the theory of *wavelet multiresolution analysis*. The application of wavelets as analysis filters for feature extraction is similar to the application of the spatial/spatial frequency representations

described in Section 3.1. However, wavelets offer potential advantages over the more classical approaches by providing an orthonormal bases that represents an image in terms of spatial scale rather than spatial frequency. Furthermore, the recent development of families of orthonormal wavelets fits within the framework of multiresolution analysis. Multiresolution analysis offers a rigorous means to study the decomposition of images into a hierarchy of resolution scales similar to the resolution pyramids and subband coding presented in Section 3.2. Wavelet multiresolution analysis may provide an elegant means to perform object recognition for problems in which the objects of interest are imaged at greatly varying distances, such as in mid-course decoy discrimination.

3.3.1 Wavelets

Wavelets are families of functions generated from a single function, known as the mother wavelet, through translations and dilations.

$$\psi^{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right)$$

The mother wavelet ψ is generally required to satisfy the condition $\int \psi(x)dx = 0$ and to decay faster than $\frac{1}{|x|}$ as $x \rightarrow \pm\infty$. Given these conditions, the function $\psi(x)$ will have some oscillations and some degree of localization. Through translation by b and dilation by a , the wavelet family for a given function $\psi(x)$ can range over a continuum of scales and cover the real line.

3.3.2 The Wavelet Transform

An appropriate family of wavelets can be used for the purpose of function decomposition and representation. The wavelet transform is the mechanism by which an arbitrary function f is decomposed and represented as a superposition of wavelets. Such a representation decomposes f into a range of components having different scales. The representation is fundamentally defined by writing the function f as integrals over a and b of $\psi^{a,b}$ with appropriate weighting coefficients. In practice, however, f is

represented by a superposition of a discrete subset of $\psi^{a,b}$. This discretization is usually written as $a = a_0^m$ and $b = nb_0 a_0^m$ with $m, n \in \mathbb{Z}$ and $a_0 > 1, b_0 > 0$. The wavelet representation of f is then written as

$$f(x) = \sum_{m,n} c_{m,n}[f] \psi_{m,n}(x)$$

with $\psi_{m,n}(x) = \psi^{a_0^m, nb_0 a_0^m}(x) = a_0^{-m/2} \psi(a_0^{-m} x - nb_0)$.

References

- [1] R. Bajcsy and L. Lieberman, "Texture Gradient as Depth Cue", Comput. Graphics image Processing, Vol. 5, pp.52-67, 1976.
- [2] A. P. Pentland, "Fractal-based description of natural scenes", IEEE Trans. Pattern Anal. Machine Intell., Vol. 17, pp. 1177-1190, 1977.
- [3] M. Porat and Y. Y. Zeevi, "The generalized Gabor scheme of image representation in biological and machine vision", IEEE Trans. Pattern Anal. Machine Intell., Vol. 10, No. 4, PP. 452-468, July 1988.
- [4] M. Clark, A. C. Bovik, and W. S. Geisler, "Texture segmentation using a class of narrowband filters", Proc Int. Conf. Acoustics, Speech, and Signal Processing, pp. 14.6.1-14.6.4, Apr. 1987.
- [5] D. Gabor, "Theory of Communication", J. IEE (London), Vol. 93, pp. 429-457, 1946.
- [6] J. Ville, "Thorie et applications de la notion de signal analytique", Cables et Transmission, Vol. 2A, pp. 61-74, 1948.
- [7] C. H. Page, "Instantaneous Power Spectra", J. Appl. Phys., Vol. 23, PP. 103-106, 1952.
- [8] L. Cohen, "Time-Frequency Distributions - A Review", Proc. IEEE, Vol. 77, No. 7, pp. 941-981, July 1989.
- [9] E. P. Wigner, "On the quantum correction for thermodynamic equilibrium", Phys. Rev., Vol. 40, pp. 749-759, 1932.

- [10] H. I. Choi and W. J. Williams, "Improved Time-Frequency Representations of Multicomponent Signals using Exponential Kernels", IEEE Trans. Acoust., Speech, Signal Proc., Vol. ASSP-37, 1989
- [11] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner Distribution - A tool for time-frequency signal analysis, Part I: Continuous-time signals", Phillips J. Res., Vol 35, No. 3, pp. 217-250, 1980.
- [12] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner Distribution - A tool for time-frequency signal analysis, Part II: Discrete-time signals", Phillips J. Res., Vol 35, No. 4/5, pp. 276-300, 1980.
- [13] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner Distribution - A tool for time-frequency signal analysis, Part III: Relations with other time-frequency signal transformations", Phillips J. Res., Vol 35, No. 6, pp. 372-389, 1980.
- [14] L. Jakobson and H. Wechsler, "A paradigm for invariant object recognition of brightness, optical flow and binocular disparity images", Pattern Recognition Lett., Vol. 1, pp. 61-68, Oct. 1982.
- [15] L. Jakobson and H. Wechsler, "Derivation of optical flow using spatiotemporal-frequency approach", Comput. Vision, Graphics Image Processing, Vol. 38, pp. 29-65, 1987.
- [16] T. R. Reed and H. Wechsler, "Tracking of non-stationarities from texture fields", Signal Processing, Vol. 8, No. 2, pp. 95-102, Jan. 1988.